

Improving the modelling of redshift-space distortions: I. A bivariate Gaussian description for the galaxy pairwise velocity distributions

Davide Bianchi^{1,2*}, Matteo Chiesa^{1,2} & Luigi Guzzo¹

¹INAF – Osservatorio Astronomico di Brera, via Emilio Bianchi 46, I-23807 Merate, Italy

²Dipartimento di Fisica, Università degli Studi di Milano, via Celoria 16, I-20133 Milano, Italy

Accepted 2014 October 3. Received 2014 October 3; in original form 2014 July 16

ABSTRACT

As a step towards a more accurate modelling of redshift-space distortions in galaxy surveys, we develop a general description of the probability distribution function of galaxy pairwise velocities within the framework of the so-called *streaming model*. For a given galaxy separation \mathbf{r} , such function can be described as a superposition of virtually infinite local distributions. We characterize these in terms of their moments and then consider the specific case in which they are Gaussian functions, each with its own mean μ and dispersion σ . Based on physical considerations, we make the further crucial assumption that these two parameters are in turn distributed according to a bivariate Gaussian, with its own mean and covariance matrix. Tests using numerical simulations explicitly show that with this compact description one can correctly model redshift-space distortions on all scales, fully capturing the overall linear and nonlinear dynamics of the galaxy flow at different separations. In particular, we naturally obtain Gaussian/exponential, skewed/unskewed distribution functions, depending on separation as observed in simulations and data. Also, the recently proposed single-Gaussian description of redshift-space distortions is included in this model as a limiting case, when the bivariate Gaussian is collapsed to a two-dimensional Dirac delta function. We also show how this description naturally allows for the Taylor expansion of $1 + \xi_S(\mathbf{s})$ around $1 + \xi_R(r)$, which leads to the Kaiser linear formula when truncated to second order, expliciting its connection with the moments of the velocity distribution functions. More work is needed, but these results indicate a very promising path to make definitive progress in our program to improve RSD estimators.

Key words: cosmology: large-scale structure of the Universe, dark energy, theory.

1 INTRODUCTION

Work on the dynamical effect known as “Redshift Space Distortions” in galaxy surveys (RSD, Kaiser 1987), has risen steadily over the past few years, following renewed interest in the context of the “dark energy” problem (Guzzo et al. 2008; Zhang et al. 2007). Produced by galaxy peculiar velocity flows that are proportional to the growth rate of structure f , RSD provide a potentially powerful way to pinpoint whether a modification of the gravity theory, rather than “dark energy”, could be the culprit of the apparent acceleration of cosmic expansion.

Estimating f (or, as it is now typical, the combination of the growth rate and the amplitude of matter clustering, $f\sigma_8$), however, requires cleaning the linear RSD bulk-flow signal from the non-linear components of the velocity field, modelling their com-

bination in a sufficiently accurate way. This is becoming more and more crucial, given the a-few-percent value of statistical errors already reachable by the available largest samples (as BOSS, Samushia et al. 2014; Reid et al. 2014) and the even more ambitious expectations for future surveys, as Euclid (Laureijs et al. 2011).

Until recently, the standard methodology to perform this modelling has been based on a modification of the linear theory formalism first derived in Fourier space by Kaiser (1987) and extended to configuration space by Hamilton (1992), (see Hamilton (1998) for a review). This “Dispersion Model”, entails convolving the linear model with an exponential damping term, to account in particular (but not only) for the evident “Fingers of God” small-scale stretching due to galaxies in groups and clusters (Peacock 1999; Peacock et al. 2001). Despite its simplicity and empirical basis, this model performs surprisingly well, as shown in a number of appli-

* E-mail: davide.bianchi@brera.inaf.it

cations (e.g. Peacock et al. 2001; Hawkins et al. 2003; Ross et al. 2007; Guzzo et al. 2008).

However, first accurate tests with simulations showed that the method produces systematic errors of up to 10% in the measured values of f , depending on the typical mass of the halos in the simulated sample (Okumura & Jing 2011; Bianchi et al. 2012). These limitations stimulated significant activity to improve RSD modelling (e.g. Tinker 2007; Percival & White 2009; White, Song & Percival 2009; Taruya, Nishimichi & Saito 2010; Reid & White 2011; Seljak & McDonald 2011; Kwan, Lewis & Linder 2012; Zhang, Pan & Zheng 2013). An overview of the different approaches is provided in de la Torre & Guzzo (2012). Several of these developments stem from the seminal paper by Scoccimarro (2004), where (among other important developments that we shall encounter later in this paper), a more general version of the dispersion model is presented. In such model the linear Kaiser description is improved by including contributions from the galaxy velocity divergence power spectrum and the velocity-density cross-power spectrum. A notable development based on the Scoccimarro form is represented by the models of Taruya, Nishimichi & Saito (2010) and their implementation in configuration space by de la Torre & Guzzo (2012).

In the same 2004 paper, Scoccimarro also discusses in general terms the so-called “Streaming Model”, which in its origins goes back to the early description of peculiar velocities by Davis & Peebles (1983). Fisher (1995) expanded this model into a more general form, which was then further generalized in the cited paper by Scoccimarro. This approach has been recently adopted for the estimator applied to the BOSS DR-9 and DR-11 data releases (Reid & White 2011; Reid et al. 2012; Samushia et al. 2014).

In the present work, we also adopt the streaming model as our working framework towards an improved description of RSD. A particularly appealing feature of this description is that it is exact, as soon as we have a complete knowledge of the (family of) Probability Distribution Functions (PDF) of galaxy pairwise velocities at any galaxy separation in the plane (r_\perp, r_\parallel) , where r_\perp and r_\parallel indicate the components of the separation perpendicular and parallel to the line of sight, respectively. Properly describing this family of PDFs is clearly the central point in the description of RSD through the streaming model. At a given separation, the corresponding PDF is in fact a pair-weighted average of all local distributions of galaxy pairs with that separation. These local distributions can in principle be completely general. In practice, they will be governed by the intrinsic properties of the galaxy flow, which are characterized in general by a bulk velocity, i.e. a mean streaming component, and a disordered component, i.e. a dispersion. This suggests that a sufficiently general description of the PDF may be possible, at any (r_\perp, r_\parallel) , in terms of the first two moments μ and σ^2 of the local distributions.

This idea is at the basis of models in which the velocity PDF is described by integrating over given functional forms, e.g. Gaussians. A first example is provided by the work of Sheth (1996), who proposes an explanation for the nearly-exponential profile of the small-scale pairwise velocity PDF. This is obtained as a weighted sum of Gaussians, where the weighting factor is related to the Press-Schechter multiplicity function and to the particle distribution within a clump.

Tinker, Weinberg & Zheng (2006) and Tinker (2007) develop a similar concept in the framework of Halo Occupation Models (HOD): redshift-space correlations are described as the sum of a one-halo and two-halo terms, reflecting respectively the cluster-

ing/virial motions of pairs of galaxies inside halos and the clustering/relative motions of pairs belonging to different halos. The one-halo term is modelled following a procedure analogue to that of Sheth, i.e. assuming that satellite galaxy velocities follow a Gaussian distribution with pairwise dispersion related to the virial dispersion of the host halo. The two-halo velocity distribution is a combination of virial motions inside halos with the halo-halo relative velocities. The latter have a distribution \mathcal{P}_h , which is assumed to be described by a superpositions of Gaussians whose mean and variance are postulated to depend on the environment (i.e. local overdensity δ) in which the two halos are found. In Juszkiewicz, Fisher & Szapudi (1998), instead, a skewed exponential distribution for the pairwise velocities is constructed in the context of Eulerian perturbation theory, based on an *ad hoc* ansatz for the pairwise velocity.

In this paper we develop the idea that the two moments of the Gaussian components behave in fact as jointly distributed random variables. More explicitly, we assume that for any given (r_\perp, r_\parallel) the overall velocity distribution can be obtained by averaging over a given family of elementary distributions \mathcal{P}_L (for example, but not necessarily, Gaussian functions) with statistical weight assigned by the joint probability distribution $\mathcal{F}(\mu, \sigma)$. We shall show that this description is general enough, as to model the redshift-space correlation function on all scales via the streaming model. Then we focus on the specific case in which \mathcal{P}_L and \mathcal{F} are respectively univariate and bivariate Gaussians, showing that even under this strong assumption the overall velocity profiles are correctly reproduced, as well as the corresponding redshift-space clustering. This simple model ultimately shows that the relevant RSD information is contained into five scale-dependent parameters, namely the mean and the covariance matrix of the bivariate Gaussian. Since the interpretation of these five parameters is clear, they can be, in principle, expressed as a function of fundamental cosmological quantities, such as the growth rate of structure. We shall explore this connection in a further work.

The paper is organized as follows. In Sec. 2 we introduce our general description of the line-of-sight pairwise velocity distribution and we discuss its implications on modelling redshift-space distortions; two specific ansatzes for the velocity PDF are discussed in detail: local Gaussianity and local Gaussianity plus global bivariate Gaussianity; in Sec. 3 we test the effectiveness of these assumptions using N-body simulations; our results are summarized in Sec. 5; we discuss perspectives for future developments and applications in Sec. 5.1.

2 DISSECTING THE STREAMING MODEL OF REDSHIFT-SPACE DISTORTIONS

The streaming model (Fisher 1995), in the more general formulation by Scoccimarro (2004), describes how the fractional excess of pairs in redshift space $1 + \xi_S(s_\perp, s_\parallel)$ is modified with respect to their real-space counterpart $1 + \xi_R(r)$:

$$1 + \xi_S(s_\perp, s_\parallel) = \int dr_\parallel [1 + \xi_R(r)] \mathcal{P}(r_\parallel - s_\parallel | \mathbf{r}). \quad (1)$$

Here $r^2 = r_\parallel^2 + r_\perp^2$ and $r_\perp = s_\perp$. This expression is exact: knowing the form of the pairwise velocity distribution function $\mathcal{P}(v_\parallel | \mathbf{r}) = \mathcal{P}(r_\parallel - s_\parallel | \mathbf{r})$ at any separation \mathbf{r} , a full mapping of real- to redshift-space correlations is provided. The knowledge of $\mathcal{P}(v_\parallel | \mathbf{r})$ at any \mathbf{r} is clearly the key point in this description. The question to be asked is how general this knowledge must be, or, in

other words, how many degrees of freedom are necessary for a sufficiently accurate description of this family of distribution functions and, as a consequence, of RSD.

The work presented in this paper stems from the attempt to answer this question. Our purpose is specifically that of finding the minimum set of physical parameters, which are still able to predict all the main features of the pairwise velocity PDFs along the line of sight. This description is required to be accurate enough, as to recover the correct redshift-space correlation function on all scales. The hope is that at a second stage such simplified, yet accurate, model can be connected to the dynamical description of RSD and applied to the data to extract information on the growth of structure (or even broader dynamical quantities characterizing gravity).

2.1 Characterizing the pairwise velocity distribution functions

Let us start with the following general points. Once a scale \mathbf{r} is fixed, the distribution function $\mathcal{P}(r_{\parallel} - s_{\parallel}|\mathbf{r})$ that enters Eq. (1) could be constructed – if we had access to galaxy velocities – by building the histogram of the relative velocities of all pairs with that separation. If we now imagine to split our Universe in sub-volumes of appropriate size, by construction we can think without loss of generality that the overall histogram of pairwise velocities [i.e. the un-normalized version of $\mathcal{P}(r_{\parallel} - s_{\parallel}|\mathbf{r})$], is given by the sum of all analogous *local* histograms. Each of the latter, once normalized, represents a specific local distribution function $\mathcal{P}_L(v_{\parallel}|\mathbf{r}, \mathbf{x}_i)$, where \mathbf{x}_i is the location of the i -th sub-volume. In principle, every $\mathcal{P}_L(v_{\parallel}|\mathbf{r}, \mathbf{x}_i)$ could be completely different. In reality, since galaxy dynamics is everywhere the result of gravitational instability and galaxy velocities in the different sub-volumes are necessarily correlated, we can reasonably assume that some fairly general, smooth parametric form is in principle able to describe the shape of all $\mathcal{P}_L[v_{\parallel}|p_1(\mathbf{r}, \mathbf{x}_i), \dots, p_N(\mathbf{r}, \mathbf{x}_i)]$, given a set of N functional parameters p_j to be determined.

Let us now imagine the global motions of galaxies within one of the sub-volumes: physically, it is reasonable to think that on a given scale the relative velocities of galaxy pairs can be characterized by the combination of a systematic, coherent component (infall onto overdensities or outflow from voids) and a random component. In other words, we are postulating that the local distribution functions can be characterized simply by their first two moments, the mean $\mu(\mathbf{r}, \mathbf{x}_i)$ and variance $\sigma^2(\mathbf{r}, \mathbf{x}_i)$. Under these conditions, for any fixed separation \mathbf{r} , we expect the values of these quantities to be a continuous function of the spatial position \mathbf{x} , and therefore described by their own distribution function (over the sub-volumes). Let us call it $\mathcal{F}(\mu, \sigma)$. Within these assumptions, the global PDF that enters Eq. (1), for a given separation \mathbf{r} , can be expressed as

$$\mathcal{P}(v_{\parallel}) = \int d\mu d\sigma \mathcal{P}_L(v_{\parallel}|\mu, \sigma) \mathcal{F}(\mu, \sigma). \quad (2)$$

The distribution function of μ and σ can be written as

$$\mathcal{F}(\mu, \sigma) \equiv \mathcal{N}^{-1} \int d^3x A(\mathbf{x}) \delta_D[\mu(\mathbf{x}) - \mu] \delta_D[\sigma(\mathbf{x}) - \sigma], \quad (3)$$

where A represents the local amplitude, i.e. the local number density of pairs¹, $\mathcal{N} = \int d^3x A(\mathbf{x})$ and δ_D are Dirac delta functions.

¹ For any given separation \mathbf{r} , we can define the number density of pairs as $A(\mathbf{x}) = [1 + \delta(\mathbf{x} - \frac{\mathbf{r}}{2})][1 + \delta(\mathbf{x} + \frac{\mathbf{r}}{2})]$ where δ is the number-density contrast. We then obtain $\mathcal{N} = 1 + \xi(r)$.

By substituting Eq. (3) into Eq. (2), we obtain

$$\begin{aligned} \mathcal{P}(v_{\parallel}) &= \mathcal{N}^{-1} \int d^3x d\mu d\sigma \mathcal{P}_L(v_{\parallel}|\mu, \sigma) \\ &\times A(\mathbf{x}) \delta_D[\mu(\mathbf{x}) - \mu] \delta_D[\sigma(\mathbf{x}) - \sigma] \\ &= \mathcal{N}^{-1} \int d^3x A(\mathbf{x}) \mathcal{P}_L[v_{\parallel}|\mu(\mathbf{x}), \sigma(\mathbf{x})], \end{aligned} \quad (4)$$

which clarifies that we are in fact modelling the global PDF as a pair-weighted mean of a (normalized) functional form parameterized by its first two moments. We define the mean of μ and σ in a compact form,

$$M_k \equiv \int d\mu d\sigma \mu^{1-k} \sigma^k \mathcal{F}(\mu, \sigma), \quad (5)$$

where $k \in \{0, 1\}$, i.e. M_0 and M_1 represent the mean of μ and σ , respectively. Similarly, we define the (tensorial) central moments,

$$\begin{aligned} C_{k_1, \dots, k_n}^{(n)} &\equiv \int d\mu d\sigma (\mu - M_0)^{n - \sum_i k_i} \times \\ &\times (\sigma - M_1)^{\sum_i k_i} \mathcal{F}(\mu, \sigma) \end{aligned} \quad (6)$$

where $k_i \in \{0, 1\}$ and $n = 0, 1, 2, \dots$ is the order of the tensor. Trivially, $C^{(0)} = 1$ and $C^{(1)} = (0, 0)$. We shall denote the moments and central moments of order n of \mathcal{P} as $m^{(n)}$ and $c^{(n)}$, respectively. Finally, we shall adopt the same notation, but adding a subscript L , to describe the moments of the \mathcal{P}_L . To ease comprehension in the development of the paper, all definitions are summarized in compact form in Table 1. From Eq. (2), it follows that

$$m^{(n)} = \langle m_L^{(n)} \rangle, \quad (7)$$

where $\langle \dots \rangle \equiv \int d\mu d\sigma \dots \mathcal{F}(\mu, \sigma)$. On the other hand

$$c^{(n)} \neq \langle c_L^{(n)} \rangle, \quad (8)$$

which implies that, for example, it is possible to obtain a skewed global distribution \mathcal{P} by superposition of non-skewed local distributions \mathcal{P}_L , as we show in more detail in the following.

2.2 The velocity distribution functions as a Gaussian family of Gaussian functions

Let us then derive first the expressions of the first few moments of the global distribution \mathcal{P} under completely general conditions². The full derivation is presented in Appendix A. The results are reported in the upper part of Table 2. A very important outcome to be noted from these computations is that even if the skewness of the local $\mathcal{P}_L(v_{\parallel}|\mu, \sigma)$ and of the distribution \mathcal{F} are negligible (i.e. we set $c_L^{(3)} = 0$ and $C_{000}^{(3)} = 3C_{011}^{(3)} = 0$, respectively), we can still obtain a global skewness, $c^{(3)} = 6M_1 C_{01}^{(2)}$, as the result of the (pair-weighted) covariance between the two moments μ and σ . This is a remarkable result, as it suggests that a simple, symmetric (i.e. unskewed) shape for \mathcal{P}_L and \mathcal{F} would be sufficient to describe without a large loss of generality the overall pairwise velocity distribution \mathcal{P} .

Let us therefore assume a Gaussian form for the local distribution functions \mathcal{P}_L , i.e.

$$\mathcal{P}_L = \mathcal{G}(v_{\parallel}|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(v_{\parallel} - \mu)^2}{2\sigma^2} \right], \quad (9)$$

² Note that in the literature, $m^{(1)}$ and $c^{(2)}$ have often been denoted with v_{12} and σ_{12}^2 , respectively, see Fisher et al. (1994).

PDF	moments	central moments
\mathcal{P}	$m^{(n)} \equiv \int dv_{\parallel} v_{\parallel}^n \mathcal{P}(v_{\parallel})$	$c^{(n)} \equiv \int dv_{\parallel} (v_{\parallel} - m^{(1)})^n \mathcal{P}(v_{\parallel})$
\mathcal{P}_L	$m_L^{(n)} \equiv \int dv_{\parallel} v_{\parallel}^n \mathcal{P}_L(v_{\parallel})$	$c_L^{(n)} \equiv \int dv_{\parallel} (v_{\parallel} - m_L^{(1)})^n \mathcal{P}_L(v_{\parallel})$
\mathcal{F}	$M_k \equiv \int d\mu d\sigma \mu^{1-k} \sigma^k \mathcal{F}(\mu, \sigma)$	$C_{k_1, \dots, k_n}^{(n)} \equiv \int d\mu d\sigma (\mu - M_0)^{n - \sum_i k_i} (\sigma - M_1)^{\sum_i k_i} \mathcal{F}(\mu, \sigma)$

Table 1. Definitions and notation adopted to describe the moments of the three probability distribution functions (PDFs) considered in this work: \mathcal{P} , \mathcal{P}_L and \mathcal{F} . n is the order of the moment and $k \in \{0, 1\}$. Throughout the text $\mu = m_L^{(1)}$ and $\sigma^2 = c_L^{(2)}$. Since we do not need to define n -th order (non-central) moments of \mathcal{F} , it is intended that $M_k = M_k^{(1)}$.

n	$m^{(n)}$	$c^{(n)}$
0	1	1
1	M_0	0
2	$M_0^2 + M_1^2 + C_{00}^{(2)} + C_{11}^{(2)}$	$M_1^2 + C_{00}^{(2)} + C_{11}^{(2)}$
3	$M_0^3 + 6M_1C_{01}^{(2)} + 3M_0(M_1^2 + C_{00}^{(2)} + C_{11}^{(2)}) + C_{000}^{(3)} + 3C_{011}^{(3)} + \langle c_L^{(3)} \rangle$	$6M_1C_{01}^{(2)} + C_{000}^{(3)} + 3C_{011}^{(3)} + \langle c_L^{(3)} \rangle$
0	1	1
1	M_0	0
2	$M_0^2 + M_1^2 + C_{00}^{(2)} + C_{11}^{(2)}$	$M_1^2 + C_{00}^{(2)} + C_{11}^{(2)}$
3	$M_0^3 + 6M_1C_{01}^{(2)} + 3M_0(M_1^2 + C_{00}^{(2)} + C_{11}^{(2)})$	$6M_1C_{01}^{(2)}$
4	...	$3(M_1^2 + C_{00}^{(2)})^2 + 6[C_{11}^{(2)}(3M_1^2 + C_{00}^{(2)}) + 2C_{01}^{(2)2}] + 9C_{11}^{(2)2}$
5	...	$60M_1C_{01}^{(2)}(M_1^2 + C_{00}^{(2)} + 3C_{11}^{(2)})$

Table 2. Expressions for the moments of the velocity distribution $\mathcal{P}(v_{\parallel})$ as a function of the moments of \mathcal{F} , in the most general case (upper panel) and under the stronger GG assumption discussed in the text. In the latter case, we also report the 4-th and 5-th central moment since the set of equations with $2 \leq n \leq 5$ can be inverted to recover C and M as a function of c and m .

such that the overall $\mathcal{P}(v_{\parallel})$ is written as

$$\mathcal{P}(v_{\parallel}) = \int d\mu d\sigma \mathcal{G}(v_{\parallel}|\mu, \sigma) \mathcal{F}(\mu, \sigma). \quad (10)$$

We shall refer to this as the “local Gaussianity” (LG) assumption.

We then assume that the bivariate distribution of the μ and σ parameters describing these Gaussians is in turn a bivariate Gaussian. This corresponds to saying that the pair-weighted distribution \mathcal{F} is given by

$$\mathcal{P}(v_{\parallel}) = \int d\mu d\sigma \mathcal{G}(v_{\parallel}|\mu, \sigma) \mathcal{B}(\mu, \sigma) \quad (11)$$

where

$$\mathcal{B}(\mu, \sigma) = \frac{1}{2\pi\sqrt{\det(C)}} \exp\left[-\frac{1}{2}\Delta^T C^{-1}\Delta\right] \quad (12)$$

and

$$\Delta = \begin{pmatrix} \mu - M_0 \\ \sigma - M_1 \end{pmatrix} \quad C = \begin{pmatrix} C_{00}^{(2)} & C_{10}^{(2)} \\ C_{01}^{(2)} & C_{11}^{(2)} \end{pmatrix}, \quad (13)$$

with $C_{10}^{(2)} = C_{01}^{(2)}$. We shall refer to this second assumption (Eq. (11)) as “Gaussian (local) Gaussianity” (GG). In the following section we shall test directly the validity of both LG and GG assumptions.

In the lower part of Table 2 we report the expressions that are obtained for the first few moments of \mathcal{P} under the GG assumptions, as discussed in the Appendix. Also on these aspects there is ample room for further developments that are not explored here. In a work in preparation we are investigating a theoretical prescription for the dependence of M and $C^{(2)}$ on \mathbf{r} ; in this framework it can also be shown that all moments can be computed up to any order through a moment generating function.

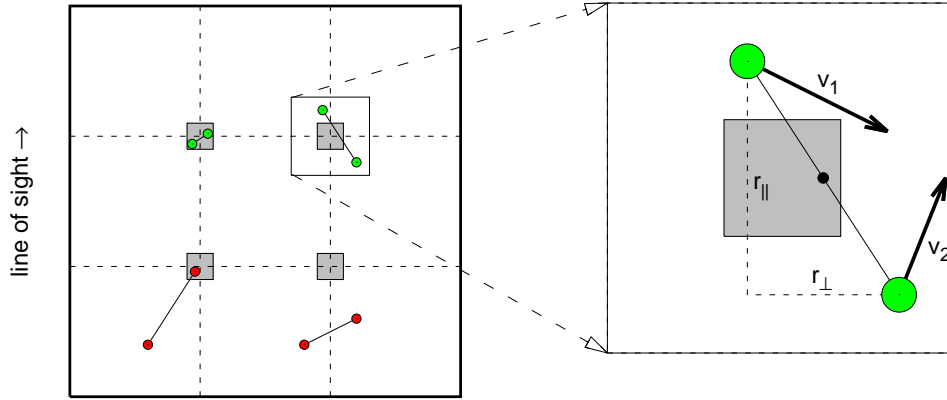


Figure 1. Two dimensional sketch of the procedure adopted to measure the local moments μ and σ^2 (i.e. $m_L^{(1)}$ and $c_L^{(2)}$) from the simulation. Gray filled squares represent $(10h^{-1}\text{Mpc})^3$ sub-volumes, ideally corresponding to different local realizations (the scale and the number of sub-volumes in the figure are arbitrary). For each local realization, only pairs with mid-point falling inside the corresponding gray area are included in the calculations (green). Pairs not satisfying this requirement are excluded from the statistics (red). For each green-like pair we measure separation (r_\perp, r_\parallel) and line-of-sight pairwise velocity v_\parallel . We then use these informations to build different local estimations (one per each sub-volume) of the velocity distribution (and its moments) as a function of the separation.

3 TESTING THE ACCURACY OF THE LG AND GG ASSUMPTIONS

To test the goodness of the LG and GG descriptions developed in the previous section, we use a properly chosen numerical simulation to directly measure the distribution of pairwise velocities at different separations. It is important to note that in this exercise we are not just checking whether the functional form of Eq. (11) is general enough to describe \mathcal{P} , for any given \mathbf{r} , by fitting for the mean M and the covariance $C^{(2)}$ of \mathcal{B} as free parameters. Rather, we want to make sure that these quantities have a well defined physical interpretation by directly measuring μ and σ from galaxy velocities in the simulation. If so, a full theoretical prediction for \mathcal{B} is in principle feasible.

For our test we use the data from the MultiDark Bolshoi run (Riebe et al. 2013), at $z = 0$. Assuming a set of cosmological parameters compatible with WMAP5 and WMAP7 data, $\{\Omega_m, \Omega_\Lambda, \Omega_b, \sigma_8, n_s\} = \{0.27, 0.73, 0.047, 0.82, 0.95\}$, this N-body simulation follows the dynamics of 2048^3 particles over a cubical volume of $(250h^{-1}\text{Mpc})^3$. If the goal were to test how accurately a given RSD model can recover the underlying cosmology (e.g. the growth rate of structure), such volume would probably be too small. For our scope, however, a small, high-resolution simulation is the best choice, as we are interested in testing in detail how the LG and GG models are capable to recover the “true” overall velocity PDF and redshift-space correlation function, given the measured local PDFs.

3.1 Practical estimate of the local velocity distribution functions

The strategy adopted to measure the local distribution \mathcal{P}_L is sketched in Fig. 1. We consider a grid with $N_L = 11^3$ nodes, which ideally correspond to N_L local realizations, i.e. the sub-volumes discussed in Sec. 2.1. N_L is basically limited by the

amount of available RAM. Since the CPU time depends essentially on the number of particles, we randomly dilute the sample down to $\approx 1.4 \times 10^7$ particles. We then store v_\parallel for all pairs whose mid-point falls inside a $10h^{-1}\text{Mpc}$ cube centered on the given $i = 1, 2, \dots, 11^3$ grid node (see Fig. 1)³. Given a separation (r_\perp, r_\parallel) , we then compute μ_i, σ_i and A_i , assuming the plane parallel approximation and rotational symmetry around the line of sight. We adopt $1h^{-1}\text{Mpc}$ bins for both separation (r_\perp, r_\parallel) and velocity v_\parallel . In order to avoid discretization effects, for any given grid node (sub-volume) and separation, the corresponding \mathcal{P}_L is included in the calculations only if sampled by more than 100 pairs. We have checked that the results reported in the following do not depend on this particular choice by repeating the procedure for different pair thresholds. From a theoretical point of view, it seems clear that any possible dependence on the threshold is mitigated by the fact that \mathcal{F} is a pair-weighted distribution. This guarantees that poorly sampled local distributions do not contribute much to the global description.

3.2 Results

In Fig. 2 we plot the results of our test for 16 representative values of the pair separation (r_\perp, r_\parallel) within the range $[0, 30] h^{-1}\text{Mpc}$. Each panel in the figure shows the measured histogram describing the pairwise velocity distribution $\mathcal{P}(v_\parallel)$ for pairs at that separation, comparing it with the two models of the same quantity constructed under the LG (blue dashed curve) and GG (red solid curve) assumptions. The panels provide an interesting overview of how the morphology of the velocity PDF can vary depending on the relative separation and orientation of the galaxy pairs, changing from exponential to Gaussian, with or without skewness. As we immediately

³ The continuous limit is readily obtained by considering a denser grid (i.e. larger N_L) with nodes surrounded by smaller cubes.

notice from the dashed and solid lines, however, this apparent complexity is in general captured quite well also under the simplifying assumptions of our models.

The agreement between the data histogram and the blue dashed line describing the LG model is indeed excellent at all separations and over the full range of pairwise velocities. This validates our first assumption, i.e. that the local distribution functions that concur to form the global one at a given separation are well described as a family of Gaussians. Also in the case of the red solid curves, corresponding to the stronger assumption that the two parameters μ and σ describing these Gaussians are also Gaussian distributed, the model PDF follows the data histogram very well. Some discrepancy is visible only at the tails of the distributions and for some separations, and is discussed below. We note that to build the PDFs under the LG assumption we have combined (i.e. summed and then normalized) all Gaussians defined by the values of (local) mean μ_i , variance σ_i^2 and amplitude A_i measured at each grid node. This is preferable to estimating the distribution of the moments \mathcal{F} by constructing a two dimensional histogram of μ and σ , as Eq. (10) would require.

Conversely, to test the GG assumption we actually estimate \mathcal{B} (i.e. its mean and covariance matrix) from the simulation and the overall distribution $\mathcal{P}(v_{\parallel})$ is then obtained by simply applying Eq.(11). The numerical estimate of \mathcal{B} inevitably adds some instability to the resulting PDFs in the GG case, which is probably at the origin of the small discrepancies observed with respect to the LG curves. The 1- and 2- σ contours of $\mathcal{B}(\mu, \sigma)$, which univocally determine $\mathcal{P}(v_{\parallel})$ under the GG assumption, are shown as insets in each panel of Fig. 2. We note that the majority of the \mathcal{B} distribution is in fact contained inside the $\sigma > 0$ plane. This is a crucial validation of the model, given that the region $\sigma \leq 0$ corresponds to an unphysical negative velocity dispersion.

The position and orientation of the elliptical contours of $\mathcal{B}(\mu, \sigma)$ vary in each panel and have important physical implications, in connection to Table 2 (lower right panel). The μ value of the centre of the ellipse corresponds to the mean of the distribution $\mathcal{P}(v_{\parallel})$. On the other hand, the variance of \mathcal{P} is related to the combination of the σ value of the centre of the ellipse with, roughly speaking, its size (as described in the third line of the GG panel of Table 2). Finally, the skewness of $\mathcal{P}(v_{\parallel})$ is due to the covariance between μ and σ , which corresponds to the rotation of the ellipse with respect to the Cartesian axes. For separations with $r_{\parallel} = 0 \ h^{-1}$ Mpc, the axes of the ellipses are substantially aligned with the Cartesian axes, and are centred at $\mu = 0$. This is expected from isotropy considerations and corresponds to zero skewness and mean of the pairwise distributions, as evident from the histograms. In particular, for the bin at separations (0, 0) the ellipse is very narrow, with very little variance on μ and large variance on σ . This is what one expects if the distribution at these separations is essentially dominated by virialized regions, for which the mean streaming is negligible.

Finally, in Fig. 3 we test how these assumptions for the velocity distribution perform when modelling redshift-space distortions through the streaming model (Eq. 1). In this figure, we compare the redshift-space correlation function $\xi_S(s_{\perp}, s_{\parallel})$ from the simulation (gray scale) to those obtained with the LG (blue dashed contours) and GG (red solid contours) models. In practice, given the real-space correlation function ξ_R , we use the streaming model to compute the $\xi_S(s_{\perp}, s_{\parallel})$ corresponding to the three different distributions \mathcal{P} previously discussed (direct measurement, LG and GG). To minimize discretization effects, we use a smooth ξ_R obtained by Fourier transforming the power spectrum given by CAMB

(Lewis, Challinor & Lasenby 2000) for the same cosmology of the Bolshoi simulation. Despite our PDFs are too poorly sampled to yield a smooth ξ_S , it is evident that the direct measurement is very accurately described by the LG approximation (blue dashed contours). As in the previous figure, the GG model (red solid contours), seems to give a slightly less stable ξ_S , nonetheless the agreement with the direct estimate remains very convincing. Note that the “unsmoothed” appearance of $\xi_S(s_{\perp}, s_{\parallel})$ is no cause for concern, as it simply reflects the limited number of “local samples” involved in the specific evaluation. A larger box together with a higher sampling over a denser grid (i.e. larger N_L) would produce a smoother function at the cost of a much heavier computational effort. However, this has no real justification for our purpose as the key point here is to test how well the dashed and solid contours reproduce the directly measured $\xi_S(s_{\perp}, s_{\parallel})$ for the same sample, including its details and deviations around the general expected form. This goal is clearly achieved here.

4 CONNECTION OF THE STREAMING MODEL WITH THE MOMENTS OF \mathcal{P}

Important insight on the role played by the moments of the overall distribution \mathcal{P} in the description of redshift-space clustering, can be gained by going back to the expression of the streaming model.

By substituting Eq.(2) in Eq.(1) and making the dependence of \mathcal{F} on \mathbf{r} explicit, we obtain

$$1 + \xi_S(s_{\perp}, s_{\parallel}) = \int d\mu d\sigma \int dr_{\parallel} [1 + \xi_R(r)] \mathcal{P}_L(r_{\parallel} - s_{\parallel} | \mu, \sigma) \mathcal{F}(\mu, \sigma | \mathbf{r}). \quad (14)$$

Let us then Taylor expand the term $(1 + \xi_R) \times \mathcal{F}$ around $r_{\parallel} = s_{\parallel}$:

$$\begin{aligned} 1 + \xi_S(s_{\perp}, s_{\parallel}) &= \sum_n \frac{1}{n!} \int d\mu d\sigma \int dr_{\parallel} (r_{\parallel} - s_{\parallel})^n \mathcal{P}_L(r_{\parallel} - s_{\parallel} | \mu, \sigma) \\ &\quad \times \frac{\partial^n}{\partial r_{\parallel}^n} \{ [1 + \xi_R(r)] \mathcal{F}(\mu, \sigma | \mathbf{r}) \} \Big|_{r_{\parallel}=s_{\parallel}} \\ &= \sum_n \frac{1}{n!} \int d\mu d\sigma m_L^{(n)}(\mu, \sigma) \frac{\partial^n}{\partial r_{\parallel}^n} \{ [1 + \xi_R(r)] \mathcal{F}(\mu, \sigma | \mathbf{r}) \} \Big|_{r_{\parallel}=s_{\parallel}} \\ &= \sum_n \frac{1}{n!} \frac{\partial^n}{\partial r_{\parallel}^n} \{ [1 + \xi_R(r)] \langle m_L^{(n)} \rangle \} \Big|_{r_{\parallel}=s_{\parallel}} \\ &= \sum_n \frac{1}{n!} \frac{\partial^n}{\partial r_{\parallel}^n} \{ [1 + \xi_R(r)] m^{(n)}(\mathbf{r}) \} \Big|_{r_{\parallel}=s_{\parallel}}. \end{aligned} \quad (15)$$

Interestingly, this is the same expansion derived by Scoccimarro (2004, Eq. 53), following a slightly different procedure.

There are a few important aspects to underline. First, we note that expression (15) does not depend on the number of moments we consider, i.e. if $\mathcal{P}_L = \mathcal{P}_L(v_{\parallel} | m_L^{(1)}, c_L^{(2)}, \dots, c_L^{(k)})$ and $\mathcal{F} = \mathcal{F}(m_L^{(1)}, c_L^{(2)}, \dots, c_L^{(k)} | \mathbf{r})$, Eq. (15) still holds for any k . Furthermore, the generic term of order n depends only on the first n local moments⁴. More in general, this expression holds if the system

⁴ This means that if, for example, $k = 2$, the first two term of the expansion do not depend on the particular functional form chosen for \mathcal{P}_L .

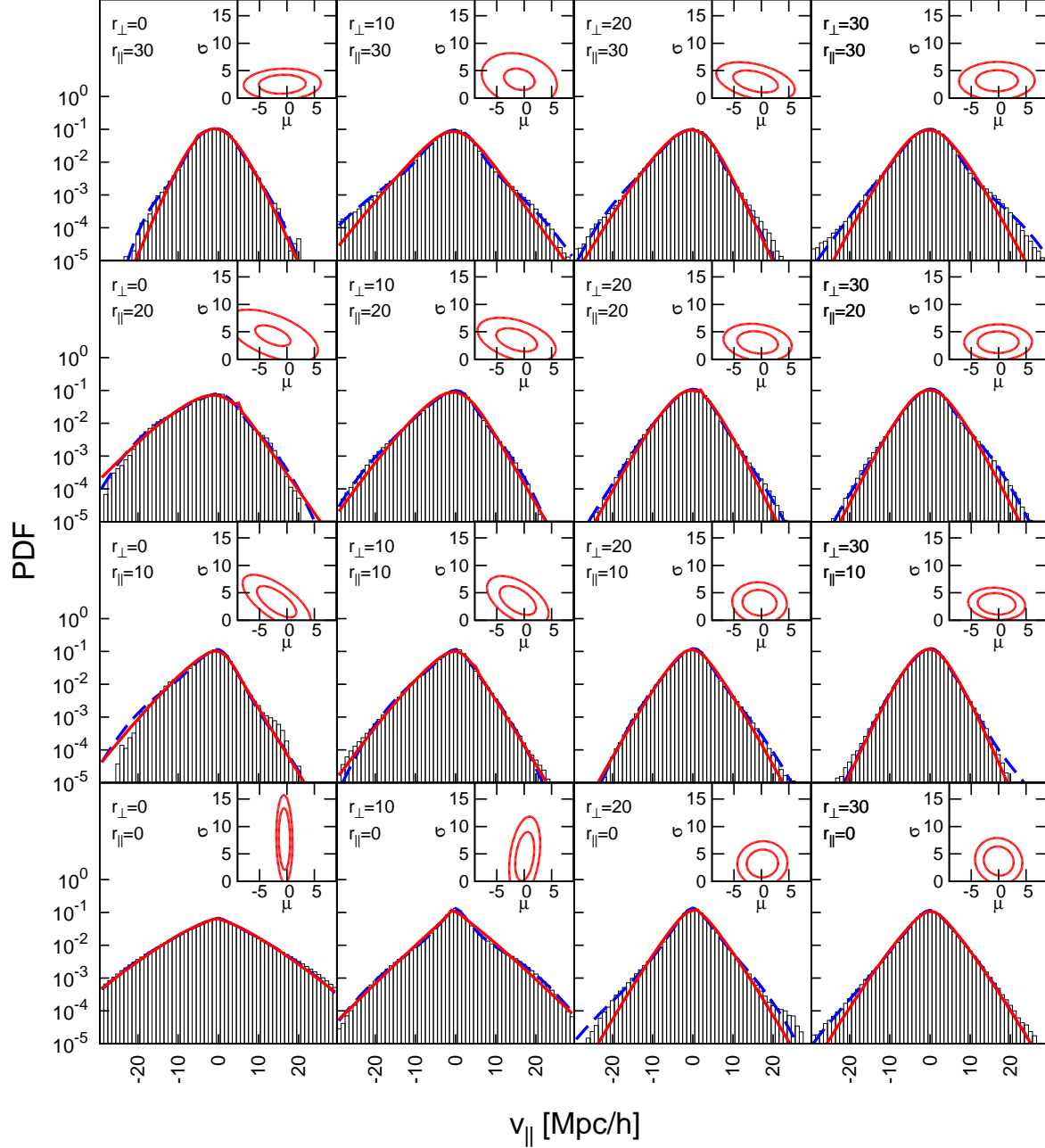


Figure 2. The distribution functions of line-of-sight pairwise velocities $\mathcal{P}(v_{\parallel})$, measured from the $z = 0$ snapshot of the simulation at a few selected separations as described in the text; the values of the separation $(r_{\perp}, r_{\parallel})$ are given in each panel in units of $h^{-1}\text{Mpc}$. The superimposed dashed and solid lines give the model PDFs for the LG and GG cases, i.e.: (a) assuming that at any separation the local distributions \mathcal{P}_L are described by Gaussian functions, for which the two moments μ_i and σ_i^2 are measured and used to empirically build the distribution function \mathcal{F} (LG, blue dashed line); (b) making the further assumption that \mathcal{F} is described by a bivariate Gaussian $\mathcal{B}(\mu, \sigma)$ as given by eq. 11 (GG, red solid line). The 1σ and 2σ contours of the $\mathcal{B}(\mu, \sigma)$ corresponding to each separation $(r_{\perp}, r_{\parallel})$ are shown in the upper right corner of each panel. Also for μ and σ units are $h^{-1}\text{Mpc}$.

can be statistically modeled in terms of a set of distributions parameterized by one or more random variables. In our specific development of the GG model in section 2.2 we have considered the case in which the random variables are $\mu = m_L^{(1)}$ and $\sigma = \sqrt{m_L^{(2)}}$. Eq. (15) can be read as a natural expansion of the redshift-space correlation function around the real-space correlation function, which corresponds to the $n = 0$ term.

Finally, if in Eq. (15) we limit the expansion to $n = 2$ and assume that $1 + \xi_R \approx 1$ and $\partial^n \xi_R / \partial r_{\parallel}^n \approx 0$ (both reasonable assumptions at large separations – but see below), we obtain

$$\xi_S(s_{\perp}, s_{\parallel}) = \xi_R(s) + \frac{\partial}{\partial s_{\parallel}} m^{(1)}(s) + \frac{1}{2} \frac{\partial^2}{\partial s_{\parallel}^2} m^{(2)}(s). \quad (16)$$

As shown by Fisher (1995) and Scoccimarro (2004), this expression corresponds to Kaiser’s linear model (Kaiser 1984). Eq. (15),

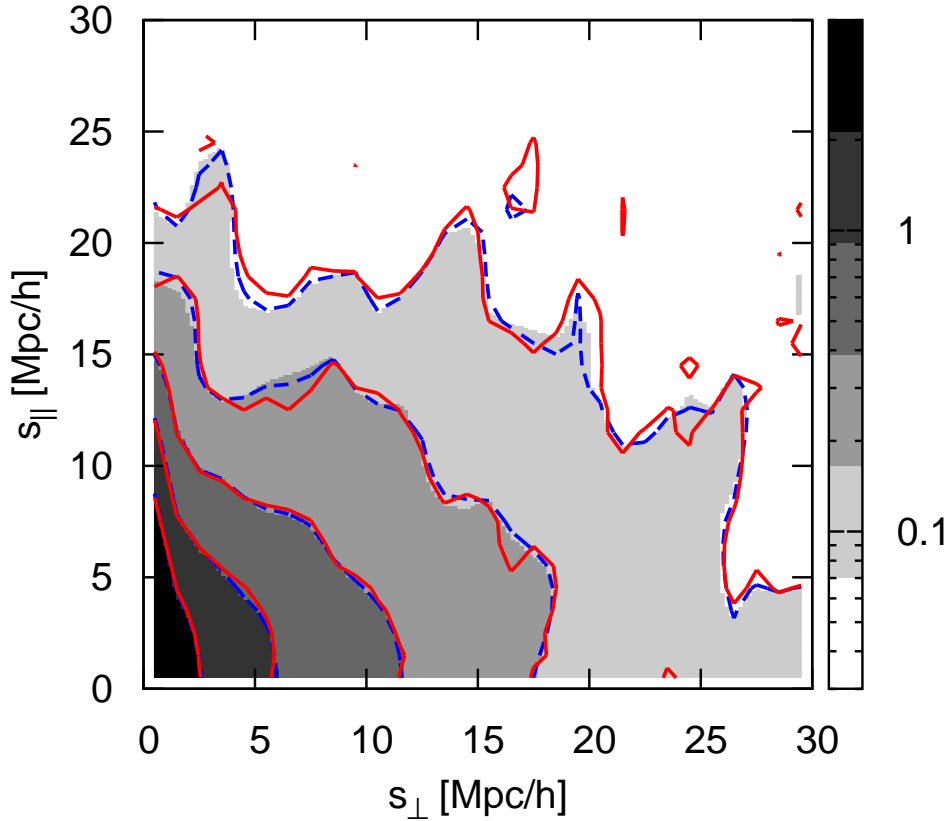


Figure 3. The redshift-space correlation function $\xi_S(s_\perp, s_\parallel)$ measured at $z = 0$ from the simulated sample as described in the text. The grayscale contours correspond to the direct measurement; the blue dashed contours correspond to fitting each local distribution of pairwise velocities \mathcal{P}_L with a Gaussian function and measuring its two moments μ_i and σ_i^2 to empirically build their distribution function \mathcal{F} ; the red solid curves are instead based on the further assumption that \mathcal{F} is described by bivariate Gaussian, as described by $\mathcal{B}(\mu, \sigma)$ in eq. 11. In practice, the contours demonstrate the impact of reducing the degrees of freedom in the form of the distribution function of pairwise velocities. The level of fidelity of the red solid contours when compared to the gray-scale ones shows the goodness of the bivariate Gaussian assumption. Note that the “unsmoothed” appearance of $\xi_S(s_\perp, s_\parallel)$ is not at all an issue, reflecting the limited number of “local samples” involved in the specific evaluation. The goal of this exercise is to show that the same $\xi_S(s_\perp, s_\parallel)$ can be obtained when using the directly measured velocity distribution, or its modelization under the increasing assumptions of the LG and GG models (see text).

therefore, naturally includes the Kaiser linear limit as a specific case. It is interesting to note that the condition that $\partial^n \xi_R / \partial r_\parallel^n \approx 0$ implies that, despite being on scales $\sim 110 \text{ h}^{-1} \text{ Mpc}$, it might be problematic to apply the Kaiser limit around the BAO peak, since there the derivative of ξ_R is far from being zero.

This derivation suggests a further interesting way to make progress in the modelling of RSD, with respect to the one developed in section 2.2. This would entail extending the proposed streaming-model expansion of Eq. (15) to some arbitrary order larger than the Kaiser ($n = 2$) limit, until the description is satisfactory.⁵ We plan to explore this approach in a future work.

⁵ Such approach requires to ensure that the expansion is perturbative, i.e. it exists a range of separations over which the $(n + 1)$ -th term is smaller than the n -th. This should be verified theoretically and against simulations.

5 SUMMARY AND DISCUSSION

Based on quite general statistical considerations, we have developed a simple model for the galaxy pairwise velocities distribution along the line of sight $\mathcal{P}(v_\parallel)$, in which, at each separation, \mathcal{P} is written as the pair-weighted mean of local distributions $\mathcal{P}_L(v_\parallel | \mathbf{p})$ (where \mathbf{p} represents an arbitrary set of parameters). More explicitly, $\mathcal{P}(v_\parallel) = \int d\mathbf{p} \mathcal{P}_L(v_\parallel | \mathbf{p}) \mathcal{F}(\mathbf{p})$, where $\mathcal{F}(\mathbf{p})$ is the overall pair-weighted joint distribution function of the parameters \mathbf{p} .

A general relation between the moments of \mathcal{P} and \mathcal{F} , is provided for the specific case in which the parameters are the pairwise velocity mean μ and standard deviation σ . We have shown that the “true” overall velocity distribution \mathcal{P} is recovered on all scales under the simple assumption that the local distributions \mathcal{P}_L are Gaussian functions whose mean μ and standard deviation σ

are in turn distributed according to a bivariate Gaussian. This corresponds in practice to compressing the whole RSD information into five well-defined physical parameters, i.e. the two central values plus three independent elements of the covariance matrix of the bivariate Gaussian. This can be seen as a natural extension to the recently proposed pure Gaussian descriptions of RSD (Reid & White 2011), which can be obtained from our model as the limiting case in which the bivariate expression becomes a two-dimensional Dirac delta.

Our approach allows the redshift-space correlation function to be expanded in terms of the individuals moments of the distribution \mathcal{P} , independently of the shape of \mathcal{P}_L and \mathcal{F} , expliciting the contribution of such moments to the description of redshift-space clustering. We have seen how this expansion recovers the Kaiser limit at large separations. Both approaches, bivariate Gaussian description and streaming-model expansion, suggest possible developments in the modelling of RSD, which we discuss below in Sec. 5.1.

We have seen in the introduction that the idea of describing the velocity PDFs by integrating over given functional forms, e.g. Gaussians, is not new in the literature. The approach presented here has similarities to previous works, but also important differences. The main novelty is in the bivariate Gaussian form, which we have shown is capable to reproduce quite well the measured velocity PDF at all separations. This includes a “natural” skewness, which is encapsulated by the covariance between the mean and variance of the family of Gaussians. More specifically, Sheth (1996) obtains the nearly-exponential profile of the small-scale pairwise velocity PDF by adding Gaussians, which are weighted by a factor related to the Press-Schechter multiplicity function and to the particle distribution within a clump. This is formally quite different from the description developed here. It is also valid for highly non-linear scales only, whereas our approach is fully general.

A comparison with the work of Tinker and collaborators (Tinker, Weinberg & Zheng 2006; Tinker 2007) is not straightforward. However, our model for \mathcal{P} could mimic their prescription for the distribution function of halo-halo relative velocities, \mathcal{P}_h , e.g. if we assumed that the dependence of μ and σ on position is totally driven by the local value of the density field. Very recently, a similar HOD approach combined with simulations has been applied to the SDSS-DR10 data, pushing the analysis to very small scales and obtaining a very precise estimate of the growth rate product $f\sigma_8$ (Reid et al. 2014).

In the work of Juszkiewicz, Fisher & Szapudi (1998), instead, the skewness of the PDF is shown to arise as a consequence of the non-trivial cross-correlation between velocity and density and some similarity with our approach is encoded in their Eq. (13), where the pairwise velocity distribution is described as a weighted sum of Gaussians. Such equation is a direct consequence of a specific choice of the PDF of the density contrast. Our approach, however, is based on completely different assumptions, as here we do not make (yet) any hypothesis on the density field nor follow any perturbative scheme.

Finally, the work of Zu & Weinberg (2013), which models the cluster-galaxy cross-correlation function in redshift space, presents some possibly interesting connections to our approach. In that paper, given a cluster-galaxy separation r_{cg} , the joint distribution $\mathcal{P}_{2D}(v_r, v_t|r_{cg})$ is considered. Here v_r and v_t are the components of the pairwise velocity parallel and perpendicular to the separation vector \mathbf{r}_{cg} , respectively. $\mathcal{P}_{2D}(v_r, v_t|r_{cg})$ is then described as the combination of a virialized term with an isotropic Gaussian velocity distribution and an infall term modelled as a skewed 2D t -distribution. The line-of-sight distribution $\mathcal{P}(v_{||}|r_{cg})$ is then ob-

tained from $\mathcal{P}_{2D}(v_r, v_t|r_{cg})$ by projection. A direct comparison with our model is not feasible, since we are dealing with auto-correlations rather than cross-correlations; still, we could imagine of adopting a similar two-dimensional approach, e.g. by modeling \mathcal{P}_{2D} as the product of two components: a full 5-parameter GG-distribution for v_r and a simplified 3-parameter GG-distribution for v_t (in the tangential direction M_0 and $C_{01}^{(2)}$ vanish due to isotropy). In practice, with this approach the angular dependence could be naturally removed at the cost of adding three more parameters, which could be an interesting test.

5.1 Perspectives

Our description of the velocity PDF opens a number of interesting questions. The definition of a full model of RSD that can be applied in practice to observational data obviously requires further developments. These include in particular building the connection to the dynamics of clustering and thus to the growth rate of structure and related cosmological parameters. While we already started working on this part of the modelling and expect results to be presented in a future paper, it is useful here to sketch some general considerations on different ways through which we expect this to be feasible.

A model-dependent possibility could be to use simulations to estimate how our bivariate Gaussian description deviates from the two-dimensional Dirac delta corresponding to the simple Gaussian model of Reid & White (2011). This would provide in practice an empirical nonlinear correction to such model. This can be seen a sort of configuration-space extension of the Fourier-space approach proposed by Kwan, Lewis & Linder (2012) and Vallinotto & Linder (2013). However, the physical meaning of the parameters in our interpretation would be completely different.

Another possibility would be to provide a theoretical prediction for the bivariate Gaussian, i.e. to derive the equations for the five parameters on which it depends (the mean values and its covariance matrix). This is in principle feasible and some analogy might be found with the approach of Seljak & McDonald (2011), in which the redshift-space density field is described in terms of the (density-weighted) velocity moments of the phase-space distribution function. No trivial relation exists between our approach and such phase-space kinetic theory. However, it is clear that in the definition of the distribution \mathcal{F} a role is played by the weighting over the number of pairs, suggesting that our description of $\mathcal{P}(v_{||})$ as a superposition of local distributions should rely ultimately on the phase-space dynamics of galaxy pairs.

More in general, rather than focusing on just extracting the growth rate of structure from RSD, we may want to consider the full information contained in the velocity PDF, if we were able to recover it from the data independently of the underlying cosmology. This would discriminate different gravity models even better. For example, it has been shown that the PDFs corresponding to the $f(R)$ gravity models introduced by Hu & Sawicki (2007) are characterized by a larger variance with respect to their Λ CDM counterparts (e.g. Fontanot et al. 2013). Clearly, in a real galaxy survey we cannot directly measure the velocity distributions. Still we could take advantage of our bivariate description to perform a Monte Carlo sampling of the five parameters on which the bivariate Gaussian depends. In essence, at each separation we are reducing the degrees of freedom by compressing a continuous function into five numbers. The acceptance/rejection criterion of the random displacements in parameter space can be obtained via the streaming-model by a standard χ^2 technique. In general, our five-parameter

compression is not enough to effectively measure the velocity PDFs on all scales (there are still too many degrees of freedom). Nonetheless, by providing an appropriate functional form for dependence of these parameters on separation, the degrees of freedom can be further reduced. Also this possibility is currently being explored for a future paper.

We can then imagine further directions of development when considering the streaming model expansion of Section 4. The first important question to be answered is how large is the range of convergence of the expansion of Eq. (15). This could be both measured from simulations and discussed theoretically, at least up to some given order and separation. If the expansion is indeed convergent over a significant range of separations, more interesting questions arise. First of all, how many velocity moments do we need to recover the “true” redshift-space correlation function on all scales? How many if we limit our analysis to some separation range, for example $s_{\perp} > 10h^{-1}\text{Mpc}$? Can we predict them? At least for the first two questions, a simple numerical approach can be easily applied.

Secondly, can we use this expansion to improve the description of how the BAO peak is distorted in redshift space? This can be explored, for example, by substituting an ad hoc functional form for the baryonic peak into Eq. (15), thus obtaining an analytic expression for the deviation from the linear Kaiser model, Eq. (16), as a function of the velocity moments. Given the precision of current and, even more, future BAO measurements from galaxy clustering, some insight into this issue might become crucial to avoid systematic effects on the peak position.

Thirdly, can we use the streaming-model expansion to directly measure the first velocity moments from the data?

Finally, we additionally note that the approach to RSD presented in this work and, more in general, all those based on the streaming model, are well suited to deal with the issue of velocity bias, since the contribution of velocity is explicit. This is an important feature in the perspective of more and more precise measurements that will require greater control over systematic effects.

ACKNOWLEDGMENTS

We thank J. Bel, S. de la Torre, V. Desjacques, J. Peacock, R. Scoccimarro and R. Sheth for useful discussions. DB and MC acknowledge financial support by the University of Milano through a PhD fellowship. This work has been developed in the framework of the “Darklight” program, supported by the European Research Council through an Advanced Research Grant to LG (Project # 291521).

REFERENCES

- Bianchi D., Guzzo L., Branchini E., Majerotto E., de la Torre S., Marulli F., Moscardini L., Angulo R. E., 2012, *MNRAS*, 427, 2420
 Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
 de la Torre S., Guzzo L., 2012, *MNRAS*, 427, 327
 Fisher K. B., 1995, *ApJ*, 448, 494
 Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J. P., 1994, *MNRAS*, 267, 927
 Fontanot F., Puchwein E., Springel V., Bianchi D., 2013, *MNRAS*, 436, 2672
 Guzzo L. et al., 2008, *Nature*, 451, 541
 Hamilton A. J. S., 1992, *ApJL*, 385, L5

- Hamilton A. J. S., 1998, in *Astrophysics and Space Science Library*, Vol. 231, *The Evolving Universe*, D. Hamilton, ed., pp. 185–+
 Hawkins E. et al., 2003, *MNRAS*, 346, 78
 Hu W., Sawicki I., 2007, *Phys. Rev. D*, 76, 064004
 Juszkiewicz R., Fisher K. B., Szapudi I., 1998, *ApJL*, 504, L1
 Kaiser N., 1984, *ApJL*, 284, L9
 Kaiser N., 1987, *MNRAS*, 227, 1
 Kwan J., Lewis G. F., Linder E. V., 2012, *ApJ*, 748, 78
 Laureijs R. et al., 2011, *ArXiv e-prints*
 Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
 Okumura T., Jing Y. P., 2011, *ApJ*, 726, 5
 Peacock J. A., 1999, *Cosmological Physics*
 Peacock J. A. et al., 2001, *Nature*, 410, 169
 Percival W. J., White M., 2009, *MNRAS*, 393, 297
 Reid B. A. et al., 2012, *MNRAS*, 426, 2719
 Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, *MNRAS*, arXiv preprint, 1404.3742
 Reid B. A., White M., 2011, *MNRAS*, 417, 1913
 Riebe K. et al., 2013, *Astronomische Nachrichten*, 334, 691
 Ross N. P. et al., 2007, *MNRAS*, 381, 573
 Samushia L. et al., 2014, *ArXiv e-prints*, 439, 3504
 Scoccimarro R., 2004, *Phys. Rev. D*, 70, 083007
 Seljak U., McDonald P., 2011, *JCAP*, 11, 39
 Sheth R. K., 1996, *MNRAS*, 279, 1310
 Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
 Tinker J. L., 2007, *MNRAS*, 374, 477
 Tinker J. L., Weinberg D. H., Zheng Z., 2006, *MNRAS*, 368, 85
 Vallinotto A., Linder E. V., 2013, *ArXiv e-prints*
 White M., Song Y., Percival W. J., 2009, *MNRAS*, 397, 1348
 Zhang P., Liguori M., Bean R., Dodelson S., 2007, *PhysRevLett*, 99
 Zhang P., Pan J., Zheng Y., 2013, *Phys. Rev. D*, 87, 063526
 Zu Y., Weinberg D. H., 2013, *MNRAS*, 431, 3319

APPENDIX A: DERIVATION OF THE MOMENTS OF THE OVERALL VELOCITY DISTRIBUTION \mathcal{P}

We sketch here the derivation of the 3-rd moment of \mathcal{P} as a function of the central moments of P_L and \mathcal{F} . We consider the most general case in which no assumptions are made on \mathcal{P}_L and \mathcal{F} (see Table 2),

$$m^{(3)} = M_0^3 + 6M_1C_{01}^{(2)} + 3M_0(M_1^2 + C_{00}^{(2)} + C_{11}^{(2)}) + C_{000}^{(3)} + 3C_{011}^{(3)} + \langle c_L^{(3)} \rangle. \quad (\text{A1})$$

All other moments can be obtained in a similar way. Under the GG assumption, it is also possible to provide the moment generating function (which will be presented in a future work), so that the moments can be computed iteratively to any order. In the following, we focus on the derivation of the most “exotic” terms of Eq. (A1), namely the correlation term $6M_1C_{01}^{(2)}$, the term contributed by the tensorial skewness $C_{000}^{(3)} + 3C_{011}^{(3)}$ and the local skewness term $\langle c_L^{(3)} \rangle$. Under the GG assumption tensorial and local skewness are set to zero by definition and the only contribution to the skewness of \mathcal{P} is given by the correlation term. The key concept in the following calculations consists of completing squares and

cubes. We have

$$\begin{aligned}
 m^{(3)} &= \int dv v^3 \mathcal{P}(v) = \int dv v^3 \langle \mathcal{P}_L(v|\mu, \sigma) \rangle \\
 &= \left\langle \int dv [(v - \mu)^3 - (-3v^2\mu + 3v\mu^2 - \mu^3)] \right. \\
 &\quad \left. \times \mathcal{P}_L(v|\mu, \sigma) \right\rangle \\
 &= \left\langle \int dv \mathcal{P}_L(v|\mu, \sigma) (v - \mu)^3 \right\rangle \quad (\text{A2})
 \end{aligned}$$

$$+ \left\langle \int dv \mathcal{P}_L(v|\mu, \sigma) 3v^2\mu \right\rangle \quad (\text{A3})$$

$$- \left\langle \int dv \mathcal{P}_L(v|\mu, \sigma) 3v\mu^2 \right\rangle \quad (\text{A4})$$

$$+ \left\langle \int dv \mathcal{P}_L(v|\mu, \sigma) \mu^3 \right\rangle. \quad (\text{A5})$$

Trivially, (A2) = $\langle c_L^{(3)} \rangle$, (A4) = $-3\langle \mu^3 \rangle$ and (A5) = $\langle \mu^3 \rangle$. This clarifies where the local-skewness term comes from. As for the (A3) summand,

$$\begin{aligned}
 (\text{A3}) &= 3 \left\langle \mu \int dv \mathcal{P}_L(v|\mu, \sigma) v^2 \right\rangle \\
 &= 3 \left\langle \mu \int dv \mathcal{P}_L(v|\mu, \sigma) [(v - \mu)^2 - (-2v\mu + \mu^2)] \right\rangle \\
 &= 3 \langle \mu(\sigma^2 + 2\mu^2 - \mu^2) \rangle \\
 &= 3 \langle \mu\sigma^2 \rangle + 3 \langle \mu^3 \rangle. \quad (\text{A6})
 \end{aligned}$$

Putting back together the summands, we get

$$m^{(3)} = \langle c_L^{(3)} \rangle + 3 \langle \mu\sigma^2 \rangle + \langle \mu^3 \rangle. \quad (\text{A7})$$

To explicit the central (tensorial) moments of \mathcal{F} , we follow a similar procedure. For example, the second summand can be written as

$$\begin{aligned}
 3 \langle \mu\sigma^2 \rangle &= 3 \langle (\mu - M_0)(\sigma - M_1)^2 \\
 &\quad - (-2\mu\sigma M_1 + \mu M_1^2 - \sigma^2 M_0 - 2\sigma M_0 M_1 - M_0 M_1^2) \rangle \\
 &= 3C_{011}^{(3)} + 6M_1 \langle \mu\sigma \rangle - 3M_1^2 \langle \mu \rangle + 3M_0 \langle \sigma^2 \rangle \\
 &\quad + 6M_0 M_1 \langle \sigma \rangle - 3M_0 M_1^2, \quad (\text{A8})
 \end{aligned}$$

which shows the origin of the $3C_{011}^{(3)}$ term. The covariance term $6M_1 C_{01}^{(2)}$ is then obtained by applying the same procedure to the second summand in the last row of Eq. (A8), namely $6M_1 \langle \mu\sigma \rangle$. Similarly, from the third summand of Eq. (A7), we recover $C_{000}^{(3)}$. In general, when developing the right hand side of Eq. (A7) polynomials in M_k and $C_{kk}^{(2)}$ are produced: putting back together all the pieces, we eventually recover Eq. (A1).